

Avaliação – Estatística Multivariada IV

Análise Discriminante Múltipla

Considere o conjunto *Auto data* do pacote ISLR do software R para desenvolver um modelo de predição para prever se um carro tem alta ou baixa quilometragem.

- (a) Crie uma variável binária, *classmpg*, que seja igual a 1 se o *mpg* for maior do que a mediana e 0, caso contrário. Você pode calcular a mediana de *mpg* no R usando a função *median()*. Note que talvez seja mais fácil usar o comando *data.frame()* para criar um conjunto de dados que contenha o *classmpg* e as demais variáveis do *Auto data*.
- (b) Explore os dados graficamente para investigar a associação entre o *classmpg* e as demais variáveis. Quais variáveis parecem ser úteis para prever o *classmpg*? Pode usar boxplots para responder a essa questão.
- (c) Divida os dados em duas amostras, uma de treino (75%) e outra de teste (25%).
- (d) Obtenha e interprete as funções discriminantes para esse estudo. Verifique também as suposições da análise. Você utilizaria a análise discriminante linear ou quadrática para classificação?
- (e) Compare a LDA e a QDA com relação à taxa de erro.
- (f) Faça uma regressão logística e avalie sua taxa de erro de acordo com alguma regra de classificação.
- (g) Faça agora uma árvore de decisão e avalie sua taxa de erro.
- (h) Utilize o método dos vizinhos mais próximos com $k = 30$ e avalie sua taxa de erro.

Data de entrega: 10/12/2018